Title: Nanopore sequencing for *Mycobacterium tuberculosis*: a critical review of the literature, new developments and future opportunities

Running title: Nanopore sequencing for *Mycobacterium tuberculosis*

Authors:

Anzaan Dippenaar[a,b,#], Sander N Goossens[a], Melanie Grobbelaar[c], Selien Oostvogels[a], Bart Cuypers[d,e], Kris Laukens[e], Conor J Meehan[b,f], Robin M Warren[c], Annelies van Rie[a]

Affiliations:

[a] Tuberculosis Omics Research Consortium, Family Medicine and Population Health, Institute of Global Health, Faculty of Medicine and Health Sciences, University of Antwerp, Antwerp, Belgium

[b] Unit of Mycobacteriology, Institute of Tropical Medicine, Antwerp, Belgium

[c] Department of Science and Innovation-National Research Foundation Centre for Excellence for Biomedical Tuberculosis Research, SAMRC Centre for Tuberculosis Research, Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Tygerberg, South Africa

[d] Department of Computer Science, University of Antwerp, Antwerp, Belgium.

[e] Molecular Parasitology Group, Institute of Tropical Medicine, Antwerp, Belgium

[f] School of Chemistry and Bioscience, Faculty of Life Science, University of Bradford, Bradford, West Yorkshire, United Kingdom

[#] Corresponding author, email address: Anzaan.Dippenaar@uantwerpen.be

Abstract:

The next-generation short-read sequencing technologies that generate comprehensive, whole-genome data with single-nucleotide resolution have already advanced tuberculosis diagnosis, treatment, surveillance and source investigation. Their high costs, tedious and lengthy processes, and large equipment remain major hurdles for research use in high tuberculosis burden countries and implementation into routine care. The portable next-generation sequencing devices developed by Oxford Nanopore Technologies (ONT) are attractive alternatives due to their long-read sequence capability, compact low-cost hardware, and continued improvements in accuracy and throughput. A systematic review of the published literature demonstrated limited uptake of ONT sequencing in tuberculosis research and clinical care. Of the 12 eligible articles presenting ONT sequencing data on at least one *Mycobacterium tuberculosis* sample, four addressed software development for long read ONT sequencing data with potential applications for *M. tuberculosis*. Only eight studies presented results of ONT sequencing of *M. tuberculosis*, of which five performed whole-genome and three did targeted sequencing. Based on these findings, we summarize the standard processes, reflect on the current limitations of ONT sequencing technology, and the research needed to overcome the main hurdles. **Summary:** The low capital cost, portable nature and continued improvement in the performance of ONT sequencing make it an attractive option for sequencing for research and clinical care, but limited data is available on its application in the tuberculosis field. Important research investment is needed to unleash the full potential of ONT sequencing for tuberculosis research and care.

Journal of Clinical Microbiology

JCM

Journal of Clinical Microbiology

Introduction

Two decades after the genome sequence of *Mycobacterium tuberculosis* (*Mtb*) H37Rv was published, sequencing technologies can now generate comprehensive genomic data with unprecedented resolution, which makes them highly attractive for research, clinical care, and applications in tuberculosis (TB) control programs(1). While the implementation of *Mtb* sequencing has been facilitated by decreases in cost, technological advances, and improved bioinformatics to translate sequence data into biologically relevant information, the initial capital expenses of Illumina sequencing platforms, sequencing reagent costs, and the need for highly trained staff remain important hurdles for wide-spread implementation, especially in high TB burden countries(2).

Long-read sequencing technologies are an enticing alternative to commonly-used short-read sequencing platforms as it allows for the analysis of complex genomic loci and large repetitive elements, both distinct characteristics of the *Mtb* genome(1). Analysing variation in these genomic regions could potentially provide a clearer understating of genes involved in host pathogen interactions and virulence. Moreover, nanopore sequencing, like PacBio single-molecule real-time (SMRT) sequencing (Pacific Biosciences, Menlo Park, California, United States), can also identify methylation status(3), which is important as epigenetic modifications in *Mtb* have been associated with drug resistance, virulence, and regulation of gene expression profiles(4, 5). Nanopore sequencing platforms developed by Oxford Nanopore Technologies (ONT), are especially attractive due to their low cost and portable hardware. The ONT MinION (Oxford Nanopore Technologies, Oxford, United Kingdom) device is capable of generating up to 30 Gigabases (Gb) of long

Journal of Clinical Microbiology

sequencing reads in 48 hours in a decentralized laboratory(6). Recent reductions in error rates, updated flow cells, lower amounts of required input DNA, and faster library preparation protocols have renewed the interest in its use in TB research and clinical applications(7-9).

We outline the unique aspects of ONT sequencing and performed a critical narrative systematic review of the published literature on ONT sequencing of *Mtb* to reflect on recent developments and future opportunities.

ONT sequencing

Nanopore sequencing is a unique, scalable technology that monitors changes in an electrical current as nucleic acids are passed through a nanopore protein. The resulting signal is decoded to provide the specific DNA or RNA sequence. The use of a nanopore for sequencing a single molecule of DNA or RNA negates the need for PCR amplification or chemical labeling of the sample for certain applications. The versatility of the platform and library preparation approaches allow for the sequencing of native nucleic acids, PCR libraries and amplified genomic targets. An overview of the general approach to the *Mtb* ONT sequencing process is shown in Figure 1. In this section, we briefly describe ONT flow cells, library preparation, base-calling and bioinformatics analysis applicable to *Mtb* ONT sequencing.

ONT Flow cell chemistry

ONT Flow cells are designed to detect current signals from k-mers as the nucleic acid molecules move through the nanopore. The early R7 flow cell chemistry associated the presence of 6-mer nucleotides with the measured current signal,

while in R9 flowcells, this has been reduced to a 3-mer. Ignoring potential base-modifications and assuming that only 4 different bases can be present corresponds to a reduction from 4096 (= 4^6) possible k-mers to 64 (4^3) possible k-mers during base-calling(10). In addition to pores that are occupied by shorter k-mers, dual reading of nucleotide sequences inside the pore was implemented in the latest R10 flow cells, implying that sequences are associated with a current signal at two different points in space and time inside the pore, resulting in an improved resolution and improved base-calling of homopolymeric regions.

Library preparation

A wide range of sequencing kits and library preparation approaches are available for ONT sequencing, each recommended for specific applications, and requiring varying quantities of input material. For example, native DNA or RNA sequences can be detected in workflows without PCR-amplification but require 400-1000 ng of input DNA(11, 12). Amplification-based approaches are recommended by ONT when input DNA is limited in quantity or quality, when control over the read length is required, and for targeted amplicon sequencing. Unless input DNA is limited, fragmentation of input DNA is not necessary, leading to a read length equal to the fragment length of the input DNA(9).

VolTRAX (Oxford Nanopore Technologies, Oxford, United Kingdom) is a portable automated sample preparation device that transforms biological samples into sequence-ready libraries and enables consistent library quality, even in the field or in the absence of elaborate laboratory infrastructure(9, 13, 14). VolTRAX is compatible with the hand-held MinION device and can multiplex up to ten samples. ONT has

also developed rapid field sequencing kits to overcome the challenges associated with cold-chain transport of reagents. Together, these developments increase the speed and simplicity of library preparation and reduce the need for specialised laboratory equipment and enhance the possibility of moving sequencing into the field and closer to point of care(15).

Base-calling

ONT-sequencers monitor the changes in electrical current as nucleic acids pass through a nanopore protein. The resulting signals are stored in FAST5 files and are decoded during the 'base-calling' process, which translates the raw signals into nucleic acid sequences in FASTQ format(9). The electrical current signal reflects the presence of a k-length nucleotide sequence (k depending on the type of flow cell) passing through the nanopore. This makes base-calling of ONT sequencing data computationally more demanding, error-prone and complex than the simple one-to-one conversion algorithm used by other sequencing technologies.

The first base-callers used hidden Markov models to estimate the likelihood that an observed signal corresponds to a particular k-mer sequence inside the nanopore(3). More recent base-callers use machine learning or hybrid models for inferring k-mers from raw signal data. Machine learning models are mainly trained on sequencing data derived from *Escherichia coli* , although several base-callers (such as Guppy(9) and Chiron(16)) allow users to *de novo* train the algorithm. Ideally, a base-caller for *Mtb* would be trained on *Mtb* sequencing data with a known ground truth (reference sequence) so that base modifications found in *Mtb* (such as m6A and m5C(17)) and species-specific sequence patterns can be captured accurately.

Bioinformatic analysis

Mtb ONT sequencing bioinformatics pipelines follow a similar approach to pipelines for short-read (e.g., Illumina) data, with quality control (read trimming), alignment to a reference genome, variant identification, and annotation to identify genomic variants for diagnosis of drug resistance or identification of transmission events(1). Some short-read pipelines (Mykrobe and TBProfiler) have already been updated to also analyse longer read sequence data(18, 19).

In contrast to most short-read pipelines which align sequence reads to the reference genome, ONT sequence analysis pipelines typically include the option to perform *de novo* genome assembly (e.g. Flye(20)). An interesting but costly approach is a hybrid assembly(21) in which the (more error-prone) long reads are used to close gaps by linking contigs and resolving repeat regions and the (accurate) short reads are mapped to the assembled contigs to correct for sequencing errors. This approach has been used for *Mtb*(21) and is especially valuable for analysis of highly repetitive regions (such as *pe/ppe* regions), for detecting structural genome variation (e.g., inversion, insertions or deletions), and for deciphering the genomes of novel *Mtb* lineages.

ONT sequencing for *Mtb* research and clinical care

We performed a search of PubMed and Scopus on 20 Jan 2021 using the search terms “tuberculosis” or “*Mycobacterium tuberculosis*” and “nanopore” or “Oxford Nanopore Technologies” or “portable sequencing” without date or language restrictions. Papers were eligible for inclusion if *Mtb* ONT data were presented for at

174 least one sample or if the development of bioinformatics tools for analysis of *Mtb*
175 ONT data was described. We identified 58 articles of which 12 were eligible. Five
176 articles focused on whole-genome sequencing (WGS)(8, 12, 21-23), three on
177 targeted sequencing(11, 24, 25) (Table 1), and four on software development for
178 long read ONT data with applications for *Mtb*(16, 18, 19, 26) (Table 2).
179
180 *Research applications*
181 The first 'proof of principle' study was published in 2016 and focussed on the
182 development of an enrichment protocol of *Mtb* DNA for ONT sequencing(22). Eckert
183 *et al.* mixed *Mtb* DNA (H37Rv and DNA from a clinical extensively drug-resistant
184 strain) with human genomic DNA (at 10% and 90% *Mtb* DNA) and used biotinylated
185 RNA baits synthesised based on *Mtb* H37Rv to capture long fragments of *Mtb* DNA.
186 They reported that unenriched mixtures resulted in very low *Mtb* genome coverage
187 while enrichment resulted in partial genome coverage. Areas with high coverage
188 depth corresponded to open-reading frames encoding transposases, which may
189 reflect redundancy of the captured sequence.
190
191 In 2018, Bainomugisa *et al*. used ONT sequencing to investigate a Beijing strain that
192 had caused outbreaks of drug-resistant TB in Papua New Guinea(21). By combining
193 a complete ONT-based genome assembly with Illumina sequencing for error
194 correction, Bainomugisa *et al.* identified all drug resistance-causing mutations, novel
195 variation, including three previously undescribed genomic deletions (1315, 1355,
196 1356 bp, respectively) and two insertions (390 and 4490 bp), multiple variants in
197 repetitive *pe/ppe* gene regions and compensatory mutations.
198

Journal of Clinical Microbiology

Clinical applications

The largest clinical study of 431 cultured isolates was published in 2020(23). Smith *et al.* aimed to validate ONT for species identification, *in silico* spoligotyping, detection of drug resistance, and phylogenetic analysis. WGS on ONT MinION showed drug resistance profiles comparable to those obtained by Illumina MiSeq (96% and 96,2% respective concordance with phenotypic drug susceptibility testing), and with equal or faster turnaround time, and competitive per sample sequencing cost (±63 USD on ONT vs 130 USD for Illumina MiSeq). Small insertions and deletions and heterozygous variants were more difficult to ascertain with high accuracy using ONT data.

A small study published in 2020 aimed to validate the ONT rapid sequencing kit for detection of drug-resistance in *Mtb*(8). Cervantes *et al.* observed that the number of reads aligned to the *Mtb* reference genome varied considerably (6,736 to 28,090) for purified DNA extracted from one laboratory and four clinical *Mtb* culture isolates. When DNA was extracted directly from two sputum specimens, the number of mapped *Mtb* reads was very low (16 and 53), and the majority of the reads produced corresponded to human DNA.

The first study of culture-free ONT sequencing of bronchioalveolar lavage and lymph node aspirates specimens was also published in 2020(12). George *et al.* achieved a mean *Mtb* genome coverage for clinical specimens ranging from 0.55x to 81x. High (99.9%) consensus accuracy from ONT data was obtained when Nanopolish, a software package designed to analyse ONT data at the signal-level, was used. Unfortunately, multiplexing resulted in insufficient genome coverage as 5 to 47% of

reads could not be reliably assigned to an input sample. Optimal ONT sequencing results were thus only achieved when one flow cell was dedicated to a single sample, rendering this approach prohibitively expensive for routine clinical settings.

Three studies assessing ONT sequencing for targeted sequencing of drug resistance loci were all published in 2020(11, 24, 25). Tafess *et al*. and Chan *et al*. used custom-made panels of 19 and 10 drug resistance-associated loci, respectively. Tafess *et al.* showed 100% agreement in drug resistance detection between ONT and Illumina MiSeq when variants with an allele frequency below 40% reported by ONT sequencing were excluded. Chan *et al.* showed 95% concordance for ONT detected variants with an allele frequency of 100% as reported by MiSeq(24, 25). The cost per sample for sequencing 19 resistance loci developed by Tafess *et al.* was 72 USD on the ONT MinION and 68 USD on Illumina MiSeq but the turnaround time was shorter using the MinION (15 hours) compared to MiSeq (38 hours)(24). Chan *et al*. reported similar per sample assay and sequencing costs on the ONT MinION device (64 USD) when sequencing 24 samples per flow cell(25). Cabibbe *et al*. assessed the GenoScreen Deeplex Myc-TB assay and found full concordance in detecting drug-resistant variants between ONT MinION and Illumina MiniSeq when applying an allele frequency threshold of 80%. The assay and sequencing costs were comparable between ONT MinION and Illumina MiniSeq, at approximately 100 Euros per sample(11).

Software development

Two papers published in 2019 focussed on the use of existing bioinformatic pipelines [Mykrobe (previously Mykrobe predictor) and TBProfiler] to process FASTQ files for

Mtb ONT WGS data(18, 19). Mykrobe predictor was one of the first software packages for species identification and *Mtb* drug resistance prediction but the use of Mykrobe predictor for ONT WGS analysis was not evaluated for *Mtb*(27). Hunt *et al.* assessed Mykrobe, the updated version which features an updated statistical model for ONT data, updated resistance library, and functionality to use a custom drug resistance library on five *Mtb* isolates. Mykrobe detected the exact same resistance causing mutations from both ONT and Illumina sequence data(19). TBProfiler, another bioinformatics tool developed to predict drug resistance and infer *Mtb* lineage and strain type from Illumina WGS data(28), was updated by Phelan *et al.* and adapted to allow the analysis of ONT data(18). TBProfiler analysis of 34 replicates of three multi-drug resistant *Mtb* isolates showed that one resistance-conferring variant (insertion in the *tlyA* gene) was missed by analysis of the ONT MinION data as compared to Illumina data(18).

Teng *et al.*, developed Chiron as an open-source base-calling algorithm. Chiron translates raw nanopore current signals directly into nucleotide sequences using deep learning neural networks(16). Chiron was trained on viral (*Escherichia virus Lambda)* and bacterial (*Escherichia coli*) sequencing reads and allows users to train the neural network in the software with their own specific genomes of interest (with distinct characteristics). When used for the base-calling of a single *Mtb* isolate sequenced with a MinION device, Chiron was shown to be more accurate than Albacore V1.1 and nearly as accurate as Albacore V2.0.1 (developed by ONT).

Finally, *Tang et al.* performed a small validation study of the MIRUReader software to identify mycobacterial interspersed repetitive unit – variable number tandem

repeat (MIRU-VNTR) typing profiles from ONT data(26). MIRUReader was able to predict the MIRU-VNTR profiles correctly from ONT MinION WGS data of 13 of the 15 *Mtb* strains assessed and the profiles were identical to those obtained using the GenoScreen MIRU-VNTR Quadruplex kit.

Critical evaluation of strengths and limitations of *Mtb* ONT sequencing

The key strengths of ONT's sequencing platforms are the low capital investment, competitive per sample sequencing cost when multiplexing, possibility of PCR-bias free library preparation, cold-chain-free sequencing reagents, fast turnaround time, the use of long reads to resolve complex genomic loci, and the ability to investigate methylation status.

The main limitation of ONT sequencing remains the suboptimal accuracy. Sequencing accuracy can be expressed as single read or consensus accuracy. In 2015, the single-read accuracy (or percentage identity of a sequence compared to its reference sequence) of ONT data was only 60%(10). The low single read accuracy was due to random read errors. Following changes in flow cell chemistry, improvements in base-calling software, development of post-sequencing correction tools, and 2D and $1D^2$ sequencing, the single read accuracy has increased to >95% for the latest R10.3 flow cells(9). While this is still lower than the 99,9% accuracy of short-read Illumina sequencing(29), accuracy data for the latest R10 flow cells (post R10.3) have not yet been published. Consensus accuracy measures the identity of a consensus sequence constructed from multiple overlapping reads originating from the same genomic location, and depends on systematic errors. For *Mtb*, the consensus accuracy for *de novo* genome assembly is estimated at 99,63% at 130x

coverage and 99,92% at 238x coverage(16, 21). As a consensus accuracy of 99,63% would correspond to >15000 errors in the 4.4Mbp *Mtb* genome, the high false positive rate of ONT sequencing still remains a barrier for TB-outbreak investigations. ONT is sensitive to errors in homopolymeric regions(10). When a stretch of identical k-mers passes the through the nanopore, a window of similar current signals is generated that complicate the determination of the number of identical nucleotides that is present. Base-calling of homopolymeric regions larger than the k-mer length recognized by the nanopore is therefore particularly challenging(30). This leads mostly to reduced ONT sequence accuracy for insertions and deletions.

The current accuracy levels achieved by ONT are likely sufficient to confidently detect drug-resistance conferring mutations(11, 18, 19, 21, 24), but may be suboptimal to detect hetero-resistance or mixed infection, and to infer transmission events. For example, where Illumina WGS data can detect 1% to 3% hetero-resistance at a depth of 400x and 100x, respectively(31), higher allele frequency thresholds (40%(24) and 80%(11)) for ONT data had to be used to achieve full concordance with Illumina sequencing for drug-resistance detection. Furthermore, detection of mixed infection in Illumina WGS data can be done accurately using QuantTB, which identifies mixed infections based on lineage-specific single nucleotide variant markers, but this tool has not been validated for ONT data.

The output and accuracy of ONT flow cells have improved over the past years. Non-specific PCR-based library preparation of sputum spiked with *M. bovis* BCG purified DNA (5%, 10%, and 15%) sequenced on R9 and R9.4 ONT flow cells showed lower

coverage bias and higher data yield using the R9.4 compared to R9 flow cells(7). In addition, the latest R10.3 ONT flow cells provide further increased throughput and have a longer signal detection area, therefore improved base-calling can be achieved. The release of the Flongle adapter in 2019 provides a low-output sequencing solution (2 Gb) at 90 USD per Flongle flow cell, which is the lowest set-up cost of any sequencing platform currently available.

Finally, an important limitation to the application of ONT sequencing in *Mtb* research, clinical care and public health lies in the limited experience to date with *Mtb* ONT sequencing, as ONT data for only 764 *Mtb* strains have been published by January 20, 2021.

Future prospects

One of the most promising recent ONT developments is the so-called “Read-Until” functionality of ONT sequencers. During Read-Until workflows, base-calling and rapid reference alignment are carried out in real-time, while the DNA or RNA molecule is passing through the nanopore(9). This sequence information is then used to decide whether a particular molecule should be sequenced further or ejected preliminarily from the pore by reversing the voltage. Thus, Read-Until allows directing more sequencing coverage towards targeted genomes or genomic regions that are of interest to the investigator. Read-Until currently reaches enrichments of 2.7x to 5.4x, as ejecting the sequencing read comes with a small, but cumulative risk of blocking the nanopore for the remainder of the sequencing run(32). Hence, the more selective one is, the more pores are blocked during the sequencing process, and the lower the Gb output of the flow cell will be. Future improvement such as

onboard nucleases that unblock blocked pores may resolve this problem. With regards to *Mtb* research and tuberculosis care, Read-Until could greatly advance the field of WGS directly from sputum or other clinical specimens as these samples typically have high amounts of human and microbial contaminant DNA compared to the low copy numbers of *Mtb* genomes(12). Additionally, Read-Until could be used to direct more sequencing coverage towards genes conferring drug-resistance in *Mtb*.

Another area of TB research where ONT sequencing could play an important role is epigenetics. The low-cost ONT hardware, development of open-source software to identify epigenetic base modifications, and long-read PCR-bias free sequencing of native DNA and RNA make ONT sequencing an attractive alternative for future epigenetic and multi-omics *Mtb* investigations(3), but such studies have not yet been performed. Studies using PacBio sequencing revealed that different methylation patterns may influence virulence, pathogenicity, and the development of (lineage-specific) drug resistance in *Mtb*(4). To date, only one study has used PacBio sequencing to combine genomic, transcriptomic, and methylation analysis of 22 *Mtb* isolates. Gomez-Gonzalez *et al.* found a relationship between DNA sequence, methylation, and RNA expression(5). Further research is needed to explore the multi-omics potential of ONT sequencing research and to verify the functional consequences of the identified mechanisms of gene expression regulation.

Conclusion

The low capital cost and portable nature of ONT hardware, the simplification and automation of sample- and library preparation steps when using VolTRAX, and continuous improvements in sequencing accuracy, suggest that ONT could have

value in mycobacteriological research laboratories, especially for detection of drug resistance. The development of the Read-Until function may accelerate researchers' ability to sequence directly from sputum samples. ONT's long-read sequencing may expand the research applications in *Mtb* sequencing beyond what is possible using short-read sequencing analysis workflows by including epigenetics and investigations of the role of the repetitive elements and complex regions of the *Mtb* genome. The lower per base accuracy levels of ONT sequencing compared to that of Illumina technologies currently limit its use in the detection of transmission events and the study of heteroresistance or mixed infections. Experiences with its application in public health, clinical care, and *Mtb* research remain limited and the lack of consensus in the bioinformatics analysis of *Mtb* ONT sequence data make its implementation in clinical care or public health laboratories premature.

Acknowledgements:

This work was supported by the Research Foundation Flanders (FWO), under grant No. G0F8316N (FWO Odysseus), the National Research Foundation (NRF), the South African Medical Research Council (SAMRC), and the Stellenbosch University Faculty of Medicine Health Sciences. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NRF or the SAMRC.

Conflict of interest statements:

The authors declare no conflicts of interest.

Author contributions: (as per applicable Journal specified categories)

Conceptualisation: AD, AVR

Validation: AD, SG, MG, SO

Writing – Original Draft Preparation: AD, SG, MG, SO, AVR

Writing – Review and Editing: AD, SG, MG, SO, BC, KL, CM, RW, AVR

Visualisation: AD, AVR

Supervision: AVR

References:

1. Meehan CJ, Goig GA, Kohl TA, Verboven L, Dippenaar A, Ezewudo M, Farhat MR, Guthrie JL, Laukens K, Miotto P, Ofori-Anyinam B, Dreyer V, Supply P, Suresh A, Utpatel C, van Soolingen D, Zhou Y, Ashton PM, Brites D, Cabibbe AM, de Jong BC, de Vos M, Menardo F, Gagneux S, Gao Q, Heupink TH, Liu Q, Loiseau C, Rigouts L, Rodwell TC, Tagliani E, Walker TM, Warren RM, Zhao Y, Zignol M, Schito M, Gardy J, Cirillo DM, Niemann S, Comas I, Van Rie A. 2019. Whole genome sequencing of Mycobacterium tuberculosis: current standards and open issues. Nat Rev Microbiol 17:533-545.
2. McNerney R, Clark TG, Campino S, Rodrigues C, Dolinger D, Smith L, Cabibbe AM, Dheda K, Schito M. 2017. Removing the bottleneck in whole genome sequencing of Mycobacterium tuberculosis for rapid drug resistance analysis: a call to action. Int J Infect Dis 56:130-135.
3. Simpson JT, Workman RE, Zuzarte PC, David M, Dursi LJ, Timp W. 2017. Detecting DNA cytosine methylation using nanopore sequencing. Nat Methods 14:407-410.
4. Phelan J, de Sessions PF, Tientcheu L, Perdigao J, Machado D, Hasan R, Hasan Z, Bergval IL, Anthony R, McNerney R, Antonio M, Portugal I, Viveiros M, Campino S, Hibberd ML, Clark TG. 2018. Methylation in Mycobacterium tuberculosis is lineage specific with associated mutations present globally. Sci Rep 8:160.
5. Gomez-Gonzalez PJ, Andreu N, Phelan JE, de Sessions PF, Glynn JR, Crampin AC, Campino S, Butcher PD, Hibberd ML, Clark TG. 2019. An integrated whole genome analysis of Mycobacterium tuberculosis reveals insights into relationship between its genome, transcriptome and methylome. Sci Rep 9:5204.
6. Jain M, Olsen HE, Paten B, Akeson M. 2016. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. Genome Biol 17:239.
7. Votintseva AA, Bradley P, Pankhurst L, Del Ojo Elias C, Loose M, Nilgiriwala K, Chatterjee A, Smith EG, Sanderson N, Walker TM, Morgan MR, Wyllie DH, Walker AS, Peto TEA, Crook DW, Iqbal Z. 2017. Same-Day Diagnostic and Surveillance Data for Tuberculosis via Whole-Genome Sequencing of Direct Respiratory Samples. J Clin Microbiol 55:1285-1298.

8. Cervantes J, Yokobori N, Hong BY. 2020. Genetic Identification and Drug-Resistance Characterization of Mycobacterium tuberculosis Using a Portable Sequencing Device. A Pilot Study. Antibiotics (Basel) 9.
9. Technologies ON. https://nanoporetech.com. Accessed 15 March.
10. Rang FJ, Kloosterman WP, de Ridder J. 2018. From squiggle to basepair: computational approaches for improving nanopore sequencing read accuracy. Genome Biol 19:90.
11. Cabibbe AM, Spitaleri A, Battaglia S, Colman RE, Suresh A, Uplekar S, Rodwell TC, Cirillo DM. 2020. Application of targeted Next Generation Sequencing assay on a portable sequencing platform for culture-free detection of drug resistant tuberculosis from clinical samples. J Clin Microbiol doi:10.1128/JCM.00632-20.
12. George S, Xu Y, Rodger G, Morgan M, Sanderson ND, Hoosdally SJ, Thulborn S, Robinson E, Rathod P, Walker AS, Peto TEA, Crook DW, Dingle KE. 2020. DNA Thermo-Protection Facilitates Whole-Genome Sequencing of Mycobacteria Direct from Clinical Samples. J Clin Microbiol 58.
13. Haan T, McDougall S, Drown DM. 2019. Complete Genome Sequence of Bacillus mycoides TH26, Isolated from a Permafrost Thaw Gradient. Microbiol Resour Announc 8.
14. Haan T, Seitz TJ, Francisco A, Glinter K, Gloger A, Kardash A, Matsui N, Reast E, Rosander K, Sonnek C, Wellman R, Drown DM. 2020. Complete Genome Sequences of Seven Strains of Pseudomonas spp. Isolated from Boreal Forest Soil in Interior Alaska. Microbiol Resour Announc 9.
15. Hoenen T, Groseth A, Rosenke K, Fischer RJ, Hoenen A, Judson SD, Martellaro C, Falzarano D, Marzi A, Squires RB, Wollenberg KR, de Wit E, Prescott J, Safronetz D, van Doremalen N, Bushmaker T, Feldmann F, McNally K, Bolay FK, Fields B, Sealy T, Rayfield M, Nichol ST, Zoon KC, Massaquoi M, Munster VJ, Feldmann H. 2016. Nanopore Sequencing as a Rapidly Deployable Ebola Outbreak Tool. Emerg Infect Dis 22:331-4.
16. Teng H, Cao MD, Hall MB, Duarte T, Wang S, Coin LJM. 2018. Chiron: translating nanopore raw signal directly into nucleotide sequence using deep learning. Gigascience 7.
17. Gong Z, Wang G, Zeng J, Stojkoska A, Huang H, Xie J. 2020. Differential DNA methylomes of clinical MDR, XDR and XXDR Mycobacterium tuberculosis isolates revealed by using single-molecule real-time sequencing. J Drug Target doi:10.1080/1061186X.2020.1797049:1-9.
18. Phelan JE, O'Sullivan DM, Machado D, Ramos J, Oppong YEA, Campino S, O'Grady J, McNerney R, Hibberd ML, Viveiros M, Huggett JF, Clark TG. 2019. Integrating informatics tools and portable sequencing technology for rapid detection of resistance to anti-tuberculous drugs. Genome Med 11:41.
19. Hunt M, Bradley P, Lapierre SG, Heys S, Thomsit M, Hall MB, Malone KM, Wintringer P, Walker TM, Cirillo DM, Comas I, Farhat MR, Fowler P, Gardy J, Ismail N, Kohl TA, Mathys V, Merker M, Niemann S, Omar SV, Sintchenko V, Smith G, van Soolingen D, Supply P, Tahseen S, Wilcox M, Arandjelovic I, Peto TEA, Crook DW, Iqbal Z. 2019. Antibiotic resistance prediction for Mycobacterium tuberculosis from genome sequence data with Mykrobe. Wellcome Open Res 4:191.
20. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol 37:540-546.

21. Bainomugisa A, Duarte T, Lavu E, Pandey S, Coulter C, Marais BJ, Coin LM. 2018. A complete high-quality MinION nanopore assembly of an extensively drug-resistant Mycobacterium tuberculosis Beijing lineage strain identifies novel variation in repetitive PE/PPE gene regions. Microb Genom 4.
22. Eckert SE, Chan JZ, Houniet D, The Pathseek C, Breuer J, Speight G. 2016. Enrichment by hybridisation of long DNA fragments for Nanopore sequencing. Microb Genom 2:e000087.
23. Smith C, Halse TA, Shea J, Modestil H, Fowler RC, Musser KA, Escuyer V, Lapierre P. 2020. Assessing Nanopore sequencing for clinical diagnostics: A comparison of NGS methods for Mycobacterium tuberculosis. J Clin Microbiol doi:10.1128/JCM.00583-20.
24. Tafess K, Ng TTL, Lao HY, Leung KSS, Tam KKG, Rajwani R, Tam STY, Ho LPK, Chu CMK, Gonzalez D, Sayada C, Ma OCK, Nega BH, Ameni G, Yam WC, Siu GKH. 2020. Targeted-Sequencing Workflows for Comprehensive Drug Resistance Profiling of Mycobacterium tuberculosis Cultures Using Two Commercial Sequencing Platforms: Comparison of Analytical and Diagnostic Performance, Turnaround Time, and Cost. Clin Chem 66:809-820.
25. Chan WS, Au CH, Chung Y, Leung HCM, Ho DN, Wong EYL, Lam TW, Chan TL, Ma ESK, Tang BSF. 2020. Rapid and economical drug resistance profiling with Nanopore MinION for clinical specimens with low bacillary burden of Mycobacterium tuberculosis. BMC Res Notes 13:444.
26. Tang CY, Ong RT. 2020. MIRUReader: MIRU-VNTR typing directly from long sequencing reads. Bioinformatics 36:1625-1626.
27. Bradley P, Gordon NC, Walker TM, Dunn L, Heys S, Huang B, Earle S, Pankhurst LJ, Anson L, de Cesare M, Piazza P, Votintseva AA, Golubchik T, Wilson DJ, Wyllie DH, Diel R, Niemann S, Feuerriegel S, Kohl TA, Ismail N, Omar SV, Smith EG, Buck D, McVean G, Walker AS, Peto TE, Crook DW, Iqbal Z. 2015. Rapid antibiotic-resistance predictions from genome sequence data for Staphylococcus aureus and Mycobacterium tuberculosis. Nat Commun 6:10063.
28. Coll F, McNerney R, Preston MD, Guerra-Assuncao JA, Warry A, Hill-Cawthorne G, Mallard K, Nair M, Miranda A, Alves A, Perdigao J, Viveiros M, Portugal I, Hasan Z, Hasan R, Glynn JR, Martin N, Pain A, Clark TG. 2015. Rapid determination of anti-tuberculosis drug resistance from whole-genome sequences. Genome Med 7:51.
29. Korlach J. 2013. A Closer Look at Accuracy in PacBio Sequencing. https://www.pacb.com/uncategorized/a-closer-look-at-accuracy-in-pacbio/. Accessed
30. Sarkozy P, Jobbágy A, Antal P. 2018. Calling Homopolymer Stretches from Raw Nanopore Reads by Analyzing k-mer Dwell Times. IFMBE Proceedings 65.
31. Dreyer V, Utpatel C, Kohl TA, Barilar I, Groschel MI, Feuerriegel S, Niemann S. 2020. Detection of low-frequency resistance-mediating SNPs in next-generation sequencing data of Mycobacterium tuberculosis complex strains with binoSNP. Sci Rep 10:7874.
32. Payne AH, Nadine; Clarke, Thomas; Munro, Rory; Debebe, Bisrat; Loose, Matthew. 2020. Nanopore adaptive sequencing for mixed samples, whole exome capture and targeted panels. BioRxiv doi:https://doi.org/10.1101/2020.02.03.926956.

Figure legend:

Figure 1. Overview of the *Mycobacterium tuberculosis* sequencing approach using an Oxford Nanopore Technologies sequencing platform. After DNA extraction, usually from cultured *Mtb* but in some cases directly from clinical specimens, ONT library preparation is done, which may include barcoding and/or PCR amplification of the sequence library. The prepared library is loaded on the flow cell inserted in the sequencer that is connected to a computer. During the ONT sequencing process, the current signal is detected and these data are stored in the FAST5 format. If live base-calling is enabled, the optional and new Read-Until function can be used to selectively sequence nucleic acid molecules of interest. Base-called sequences are stored in the FASTQ format, which is analysed bioinformatically. *Mtb* ONT sequencing has applications in fundamental research, clinical care and public health. Abbreviations: *Mtb*: *Mycobacterium tuberculosis*, ONT: Oxford Nanopore Technologies, PCR: polymerase chain reaction.

547 **Tables:**

548 **Table 1. Publications using Oxford Nanopore Technologies sequencing data for *Mycobacterium tuberculosis***

First author	Year	Ref	WGS or targeted	Type of strain, cultured isolate or specimen	N*	Sample preparation details and library preparation kit	Device	Flow cell used	Base-calling	Bioinformatic analysis	Main study aim	Main study findings
Eckert	2016	(22)	WGS	Laboratory (*Mtb* H37Rv), culture isolate	1	Biotinylated RNA bait enrichment, SQK-MAP003 or SQK-MAP004	MinION	Not listed	Metrichor 2D	MinKNOW, Poretools, BLASR, LAST	Evaluate an adapted DNA enrichment protocol for MinION sequencing	DNA enrichment resulted in partial *Mtb* genome coverage
				Clinical, culture isolate	1							
Bainomugisa	2018	(21)	WGS	Clinical, cultured isolate	1	SQK-LSK108	MinION	R9.4	Albacore	MinKNOW, Nanopolish, Racon, Pilon, MUMmer, Canu, Circulator	Use NS plus short-read sequencing to assemble the an XDR *Mtb* genome	Identification of known and novel genomic variants
Smith	2020	(23)	WGS	Clinical, cultured isolates	431	SQK-LSK109	MinION	R9.4	Guppy with FlipFlop Fast algorithm	QCAT, Minimap2, BWA mem, SAMTools, Kraken	Assess ONT sequencing for species identification, *in silico* spoligotyping, resistance prediction and phylogenetics	Performance and cost of ONT is comparable, to Illumina for genotyping and detection of resistance
Cervantes	2020	*(8)*	WGS	Laboratory (*Mtb* HN878), cultured isolate	1	Rapid Sequencing Kit	MinION	R9.4	Albacore	EPI2ME, What's In My Pot, antimicrobial resistance mapping	Evaluate ONT for WGS for drug resistance prediction from cultured and	Number of *Mtb* reads varied considerably and was very low for 2 when DNA was

				Clinical, 4 cultured isolates, 2 specimens	6					application	uncultured *Mtb*	extracted directly from sputum
George	2020	(12)	WGS	Laboratory (*M. bovis* BCG), cultured isolate (used to spiked sputum)	1#	SQK-LSK109	GridION	R9.4	Guppy	Porechop, Centrifuge, in-house CRuMPIT workflow, Minimap2, SAMTools, Pysam	Develop an undemanding, cost-effective method for sequencing *Mtb* directly from clinical specimens	Use of a low-cost thermo-protection buffer and a single flow cell per sample resulted in sufficient *Mtb* genome coverage
				Clinical, specimens	20							
Tafess	2020	(24)	Targeted	Clinical, cultured isolates	163	PCR amplification of 19 loci, SQK-LSK108	MinION	R9.4	Albacore	BacterioChek-TB, BWA	Develop targeted-sequencing for Illumina MiSeq and ONT for prediction of resistance.	100% concordance between ONT and Illumina when low frequency variants are excluded
Chan	2020	(25)	Targeted	Clinical, specimens	12	PCR amplification of 10 loci, Ligation Sequencing 1D kit	MinION	R9	MinKNOW	Porechop, Minimap2, Nanopolish, Qualimap	Develop targeted-sequencing workflows for Illumina MiSeq and ONT for prediction of resistance	95% concordance between ONT and Illumina for fixed variants
Cabibbe	2020	(11)	Targeted	Clinical, specimens	104	Deeplex Myc-TB PCR amplification, SQK-LSK108	MinION	R9.4	Albacore	Guppy, Porechop, Minimap2, SAMTools, VarScan2, NanoPack, AlignQC, Qualimap2	To evaluate the compatibility of Deeplex Myc-TB, with ONT MinION.	ONT MinION and Illumina MiniSeq results were fully concordant for drug resistance prediction.

*N refers to the number of sequenced samples. #20 replicates of one sample. Abbreviations: WGS: whole genome sequencing; BCG: Bacillus Calmette-Guérin; ONT: Oxford Nanopore Technologies; PCR: polymerase chain reaction; XDR: extensively drug resistant; *Mtb*: *Mycobacterium tuberculosis*

552 **Table 2. Published software for analysis of *Mycobacterium tuberculosis***

553 **Oxford Nanopore Technologies sequence data**

First author	Year	Reference	Name	Purpose	Number of samples
Hunt	2019	(19)	Mykrobe	Drug resistance prediction, species identification	5 *Mtb* clinical isolates
Phelan	2019	(18)	TBProfiler	Drug resistance prediction, *Mtb* lineage assignment	34 replicates of 3 *Mtb* clinical isolates
Teng	2018	(16)	Chiron	ONT sequencing base-caller	1 *Mtb* clinical isolate
Tang	2020	(26)	MIRUReader	*In silico* MIRU-VNTR from long-read *Mtb* sequencing data	15 *Mtb* clinical isolates

554 All publications listed used *Mtb* WGS generated using an ONT MinION device.

555 Abbreviations: *Mtb*: *Mycobacterium tuberculosis;* ONT: Oxford Nanopore

556 Technologies; MIRU-VNTR: mycobacterial interspersed repetitive unit - variable

557 number tandem repeat.

558